# Python 大众点评数据分析

## 1：提出问题

随着移动互联网的发展，使得越来越多的人通过软件选择去哪就餐。大众点评作为领先的本地生活消费平台、独立第三方消费点评网站，致力于为顾客提供餐饮领域的商户信息并且发布消费评价，具有较全面的餐饮信息。

问题1：菜系品种与店家评分是否存在相关关系？

问题 2：店家的得分呈现什么分布？

问题 3：菜系与口味，环境，服务的得分关系。

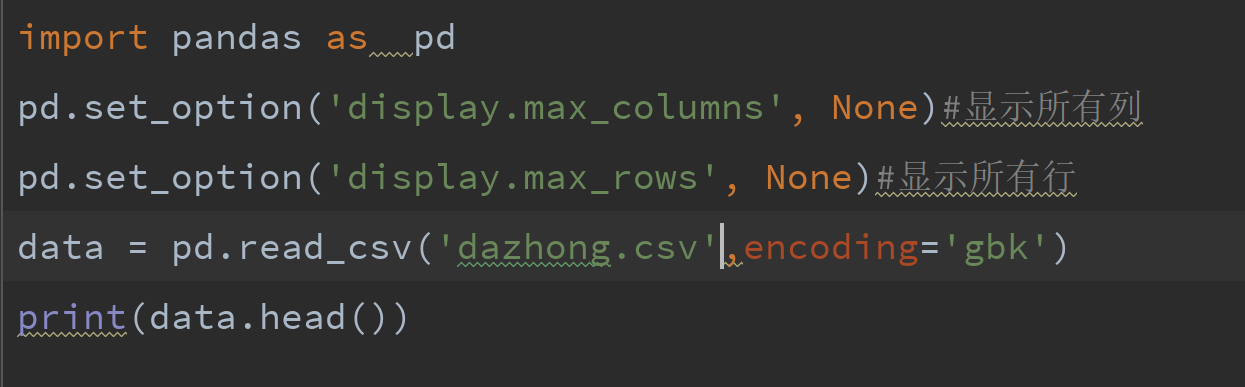
问题 4：粤菜店家的总评与其口味，环境，服务得分的关系。

## 2：理解数据

### 2.1 采集数据

本报告使用的数据为大众点评网广州站美食分类下经营不同种餐品的店家评价和评论信息。数据集共包含3124条餐厅数据，按经营餐品分类，其中粤菜722家、粥粉面572家、川菜566家、日式料理595家、西餐669家。数据格式为csv。

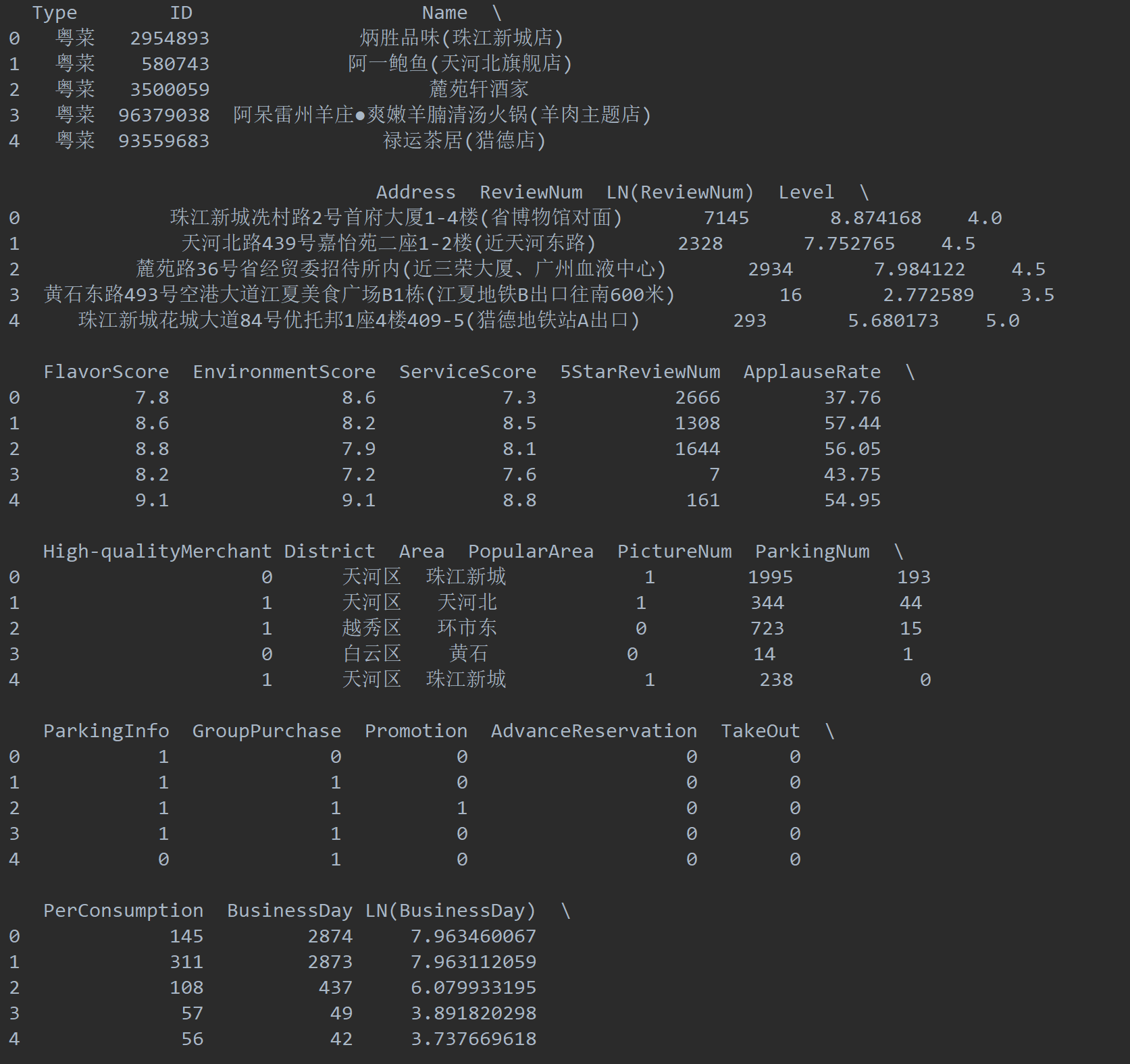
### 2.2 导入数据

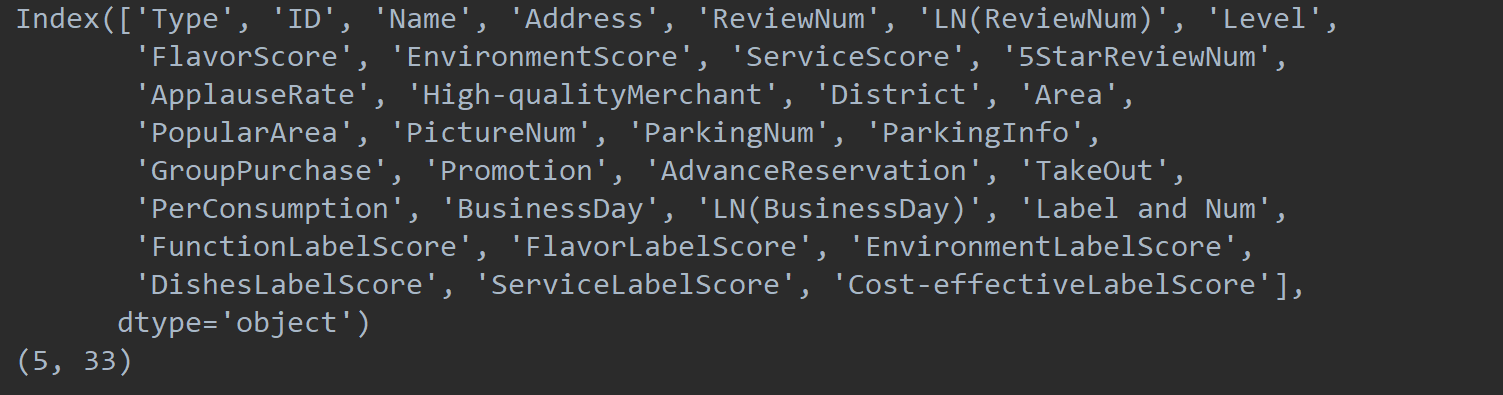


本报告使用Pandas库，读取CSV数据。

### 2.2 数据总览





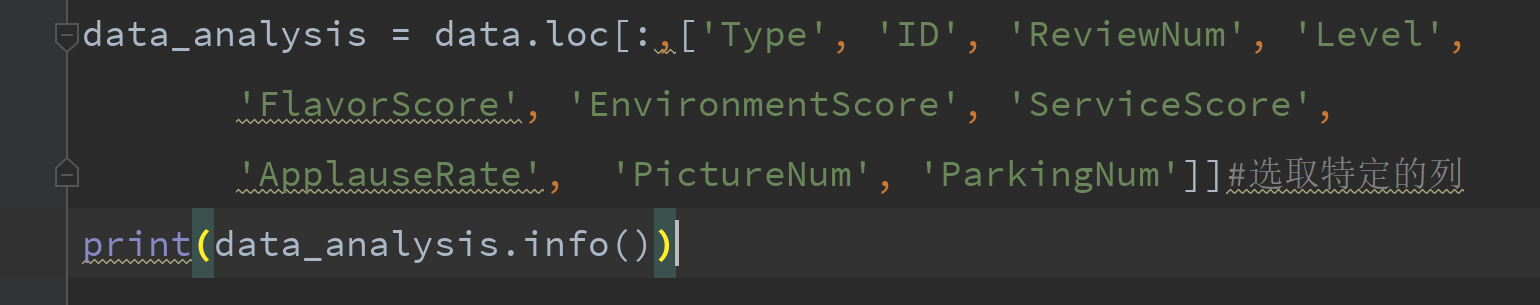


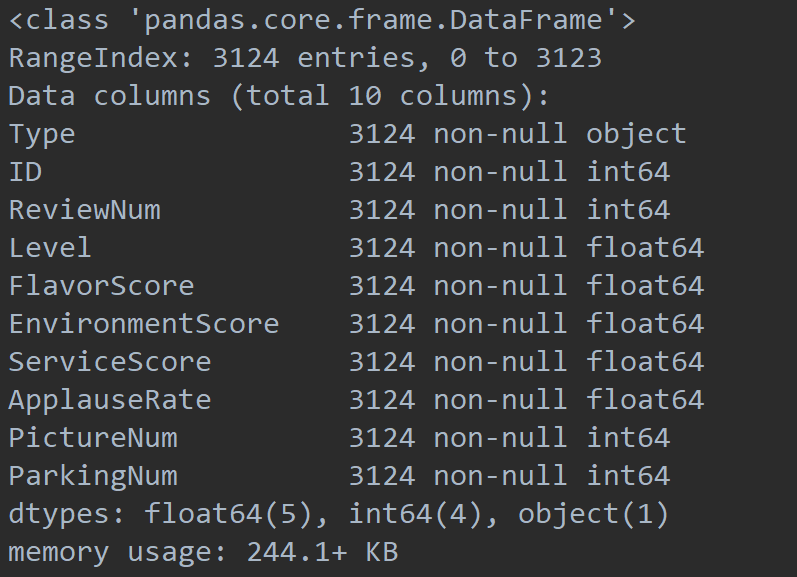
由上图可知，本数据集共有33列。因为采用head的方式进行总览数据信息，图片中仅显示了前五行信息。

## 3数据清洗

本报告将研究菜肴类型，商家等级，回头率，味道，环境，服务，好评率，以及店家展示的图片数量和停车位数量 之间的关系。

### 3.1 建立目标数据



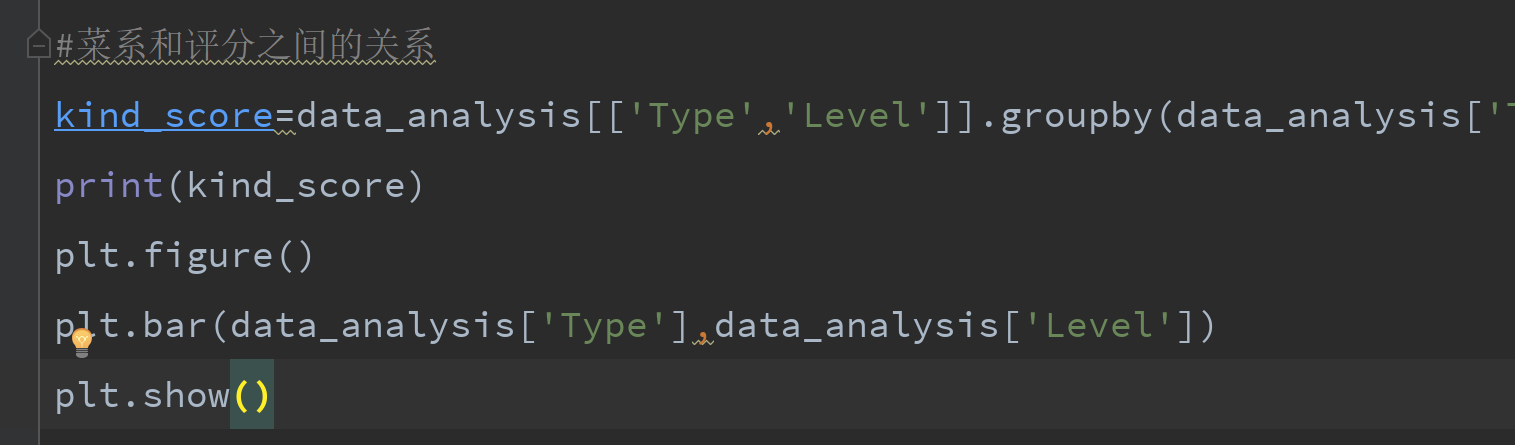


由此可知，目标数据共3124行，无缺失数据。其中数据类型有float，int和object。其中字段的含义为:

|  |  |
| --- | --- |
| Type | 类型 |
| ID | 店家编号 |
| ReviewNum | 回头课数量 |
| Level | 店家等级 |
| Flavorscore | 口味得分 |
| EnvironmentScore | 环境得分 |
| ServiceScore | 服务得分 |
| ApplauesRate | 好评率 |
| PictureNum | 图片数量 |
| ParkingNum | 停车位数量 |

## 4 数据分析及可视化

### 4.1 店家菜系与店家得分



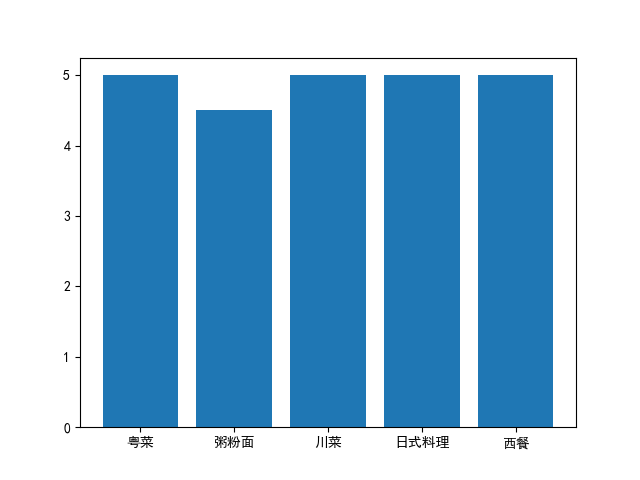
川菜 3.727032

日式料理 3.902521

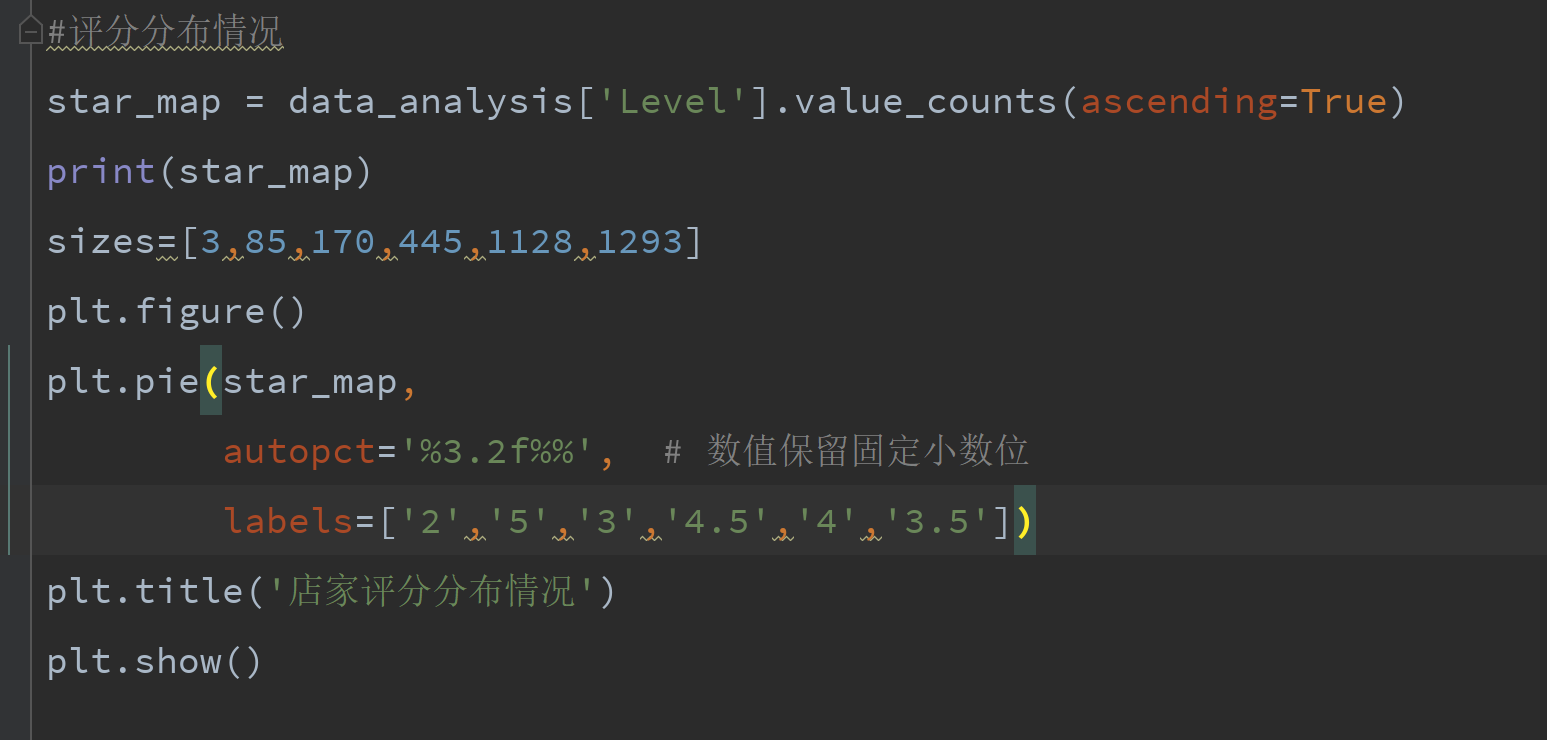
粤菜 3.918283

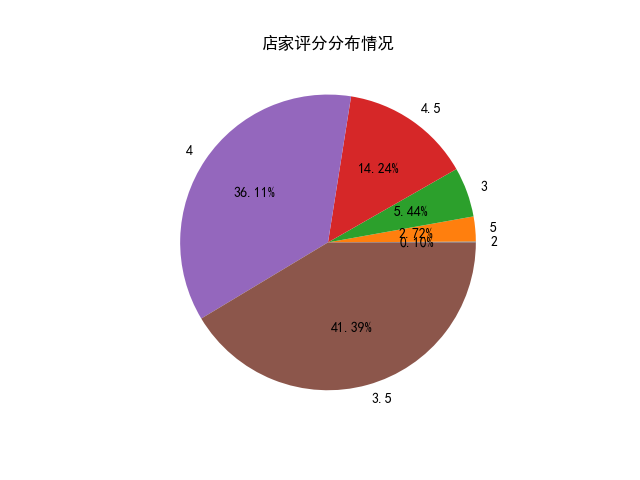
粥粉面 3.480769

西餐 4.079970



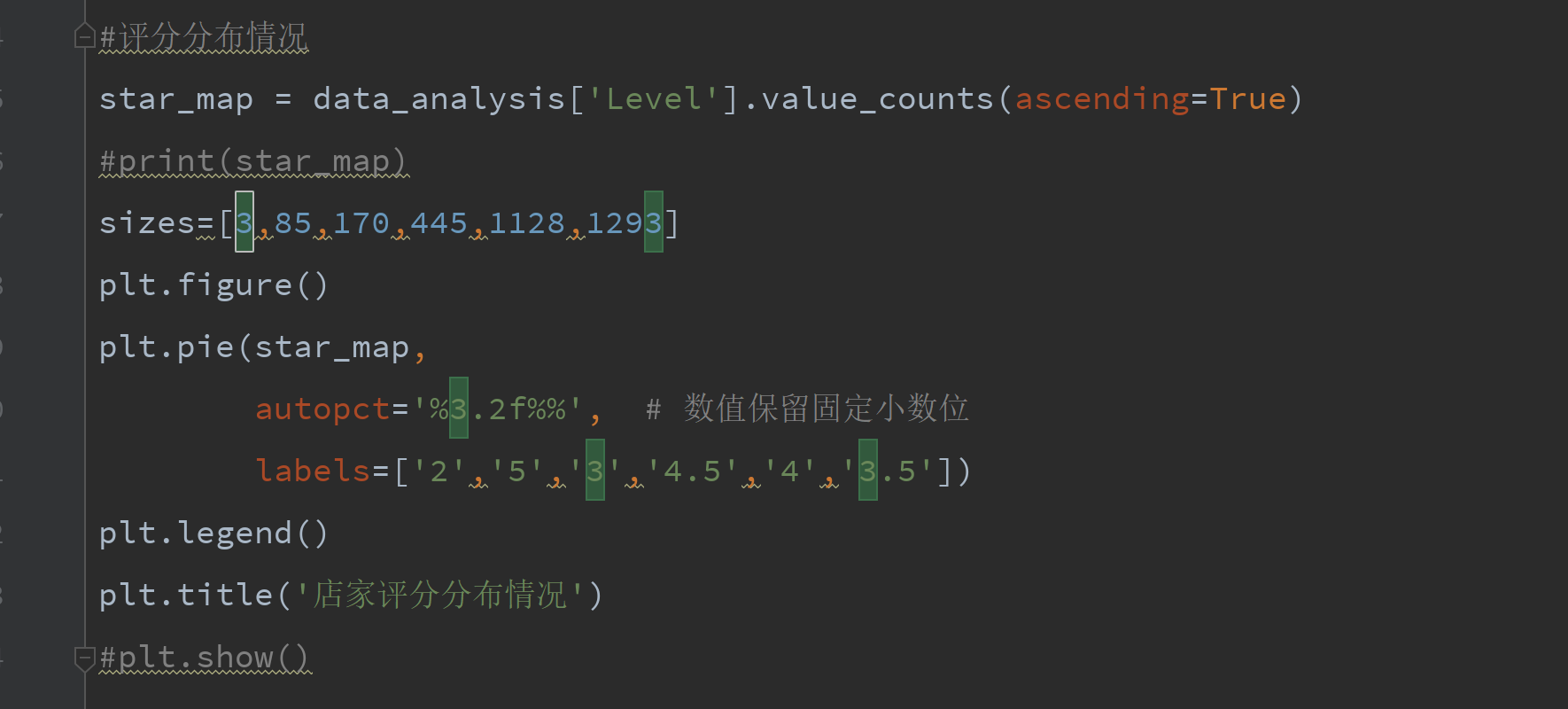
### 4.2 店家等级分布情况





由上图可知，在所有店家中，4分以上的店家超过半数。其中3.5-4分占据超过3/4的商店总数。

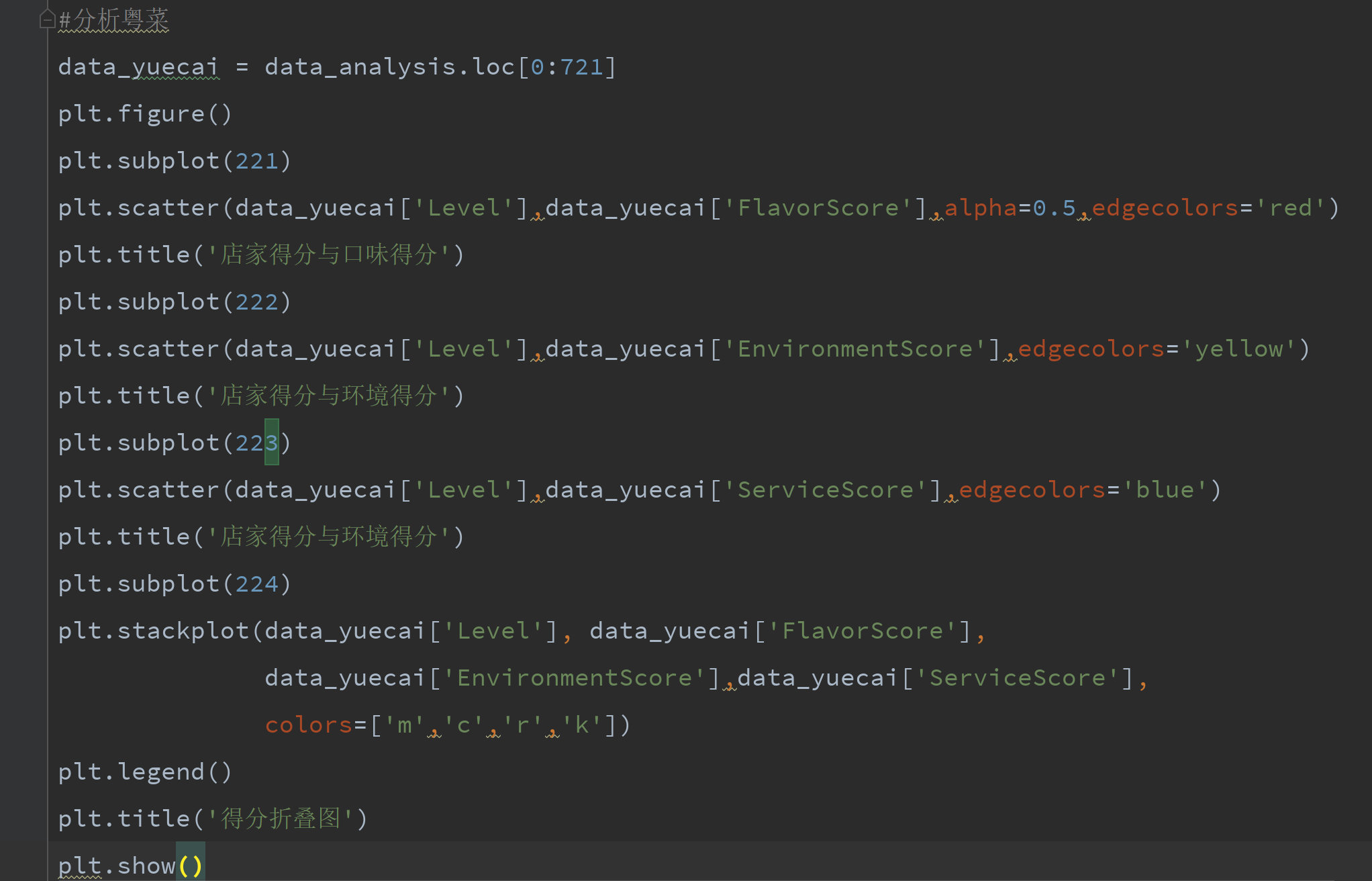
### 4.3关于店家菜系，口味，环境，服务得分分布情况



图片包含 文字, 地图

已生成极高可信度的说明

### 4.4 粤菜店家 情况分析



图片包含 屏幕截图

已生成极高可信度的说明

### 4.5 评价词云



根据客户的评价，建立评价词云。由词云可以看出，顾客在用餐时最为关注的是 环境，服务，以及口味情况。其中用餐环境是顾客最为关注的部分。

参考文献：

<http://opendata.pku.edu.cn/dataset.xhtml?persistentId=doi:10.18170/DVN/EB6KJ1>